DOCUMENT RESUME

ED 449 680 FL 026 582

AUTHOR Nakamura, Yuji
TITLE Measuring Speaking Skills through Multidimensional Performance Tests.
PUB DATE 1999-03-00
NOTE 17p.; A version of this paper was resented at the 20th Language testing Research Colloquium (20th, Monterey, CA, March 9-12, 1998). Published by International Christian University.
PUB TYPE Journal Articles (080) -- Reports - Research (143) -- Speeches/Meeting Papers (150)
LANGUAGE English, Japanese
JOURNAL CIT Educational Studies; v41 p99-113 Mar 1999
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; Communicative Competence (Languages); *English (Second Language); Foreign Countries; Higher Education; Interviews; *Language Proficiency; *Language Tests; Language Usage; *Oral Language; Second Language Instruction; Second Language Learning; Sociolinguistics; Test Validity
IDENTIFIERS Japan

ABSTRACT

This paper investigates the effectiveness of students' real communication capabilities by investigating the target language use in monologue, dialogue, and conversational "multilogue" situations. It explores Japanese college students' oral English proficiency by focusing on the linguistic and pragmatic aspects in six types of teaching tests: (1) one way monologue: speech-making overall test; (2) monologue/semi-dialogue: tape-mediated sociolinguistic test; (3) dialogue: face-to-face interview test; (4) dialogue/semi-multilogue: group interview test; (5) small-size multilogue: small-group discussion test; and (6) large-size multilogue: large-group discussion. The results demonstrate that some students who are good at monologue tests or even dialogue tests are not good at larger discussion-oriented or group activity tests. Students should be urged to develop oral communication-oriented learning habits as fluent oral communication in the target language is the ultimate goal for most if not all students. It is uncertain if a lesser ability to communicate in multilogue settings is related to Japanese culture or to some other variable. (Contains 12 references.) (KFT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

国際基督教大学学報 I－A
「教育研究 41」　抜 刷
1999 年 3 月 31 日 発 行

ED 449 680

Measuring Speaking Skills through Multidimensional Performance Tests

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Yuji Nakamura

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

International Christian University

Yuji Nakamura

BEST COPY AVAILABLE

『教育研究』41（1999）pp.99－113
Educational Studies, 41（1999）pp.99－113
International Christian University

Measuring Speaking Skills through Multidimensional Performance Tests

Tokyo Keizai University Yuji Nakamura

1. Theoretical background and rationale

One of the major objectives of teaching oral communication is enhancing students ability to use oral language in various sociolinguistic contexts. Even if we limit the contexts to academic settings, there are situations where students perform sociolinguistically appropriate to the context of spoken language. Asking questions in lectures is different from participating in discussions or making oral presentations before peers. Students are required to perform appropriately with spoken English in each context. We need to assess their real communication ability by investigating the target language use domains in which students will be required to appropriately perform in oral language.

Some students, who are good at monologue type tests, even dialogue type tests, cannot perform satisfactorily in the discussion test/ activities regardless of their outgoing personality which is generally considered a positive factor in speaking. This may be caused by the lack of discussion skill in Japanese, by the classroom circumstances where students are hesitant to speak out, or a national characteristic. Needless to say, it is worthwhile to investigate the cause of students performance differences. However, a great variety of tests for communication

ability give positive washback to students and urges them to realize the framework of communication.

2. Purpose of the research

This paper explores Japanese college students oral proficiency, by focusing on the linguistic and pragmatic aspects, from three dimensions : Monologue, Dialogue and Multilogue* (cf. Crystal 1995), and examines the specificity of their speaking ability in one test and the generalizability of their speaking ability among different tests.

Note:

Crystal (1995) states that a factor which fundamentally influences the linguistic character of a use of language is the number of participants involved in the activity. He first introduces two theoretically clear-cut distinctions: monologue, in which only one person is involved in the linguistic act, and dialogue in which two people are involved. In addition, he proposes the third term multilogue to explain the complexity of dialogue (whether the act is a pure dialogue, or a conversation as a series of mini-dialogues, or unison prayer at church). The three terms used in this present research are based on his idea, although the detailed content of each will not necessarily be the same as his idea because this research will only focus on the speaking skill as explained later.

3. Research design and methods

The linguistic aspect includes grammar, vocabulary, discourse etc. while

the pragmatic aspect deals with the level of sociolinguistic appropriateness in spoken utterances. Three dimensions are mainly categorized based on the number of people involved in the context or test situations as follows: 1) One-Way Monologue such as a speech making test or a visual description test, 2) Two-Way Dialogue such as a live interview test or a tape mediated speaking test, and 3) Multilogue such as a small group(3 - 4 people) discussion test or a large group (more than 20 people) discussion test.

Forty-six Japanese college students took these six different speaking tests. The obtained results will be examined mainly through the correlations among the six components.

3. 1. Six Types of Speaking Tests

1) One Way Monologue: Speech Making Overall Test (SO)

Please choose one topic you want to talk about among the ten topics given below and give a two-minute speech about it.

(my friends, my family, part-time work, my hobbies, traveling, fashion, college life, sports, e-mail, internet)

2) Monologue/ Semi-Dialogue (monologue in the sense that listening is not involved, and semi-dialogue in the sense that a student speaks to a virtual listener in a certain context on tape): Tape Mediated Sociolinguistic (TMS) Test

Please give an appropriate response in English in each context.

Context 1 (Apologizing and making an excuse)

You are late for your class. You missed the school bus. Please apologize and make an excuse to your teacher.

Context 2 (Complaining and requesting)

You are in a non-smoking section of a waiting room at the airport. Someone started smoking. You have a cold and a sore throat. Please complain about the smoking and request him/her to stop it.

Context 3 (Asking for repetition)

You didn't understand what your teacher said. You want the teacher to repeat it. Please make a request to your teacher.

Context 4 (Interrupting)

Your supervisor is working in his office. You want to interrupt him for a moment to talk with him. What do you say ?

Context 5 (Telephoning; taking a message)

You answer the phone. Someone wants to talk with your father. But he is out now. What will you say ?

3) Dialogue: Face to Face Interview (FFI) Test

This is a 5-minute conversation with the classroom teacher following the suggested procedure below.)

1. greeting

self-introduction (please introduce yourself by giving your name, the name of your high school and your hobbies)

2. short answer interview

Where do you come from?

Do you belong to a club?

Do you have any brothers or sisters?

How do you come to school?

How long does it take to come to school?

Do you have a part-time job?

3. long answer interview

Can you tell me something about your high school?

How many students are there?

Where is it located?

Can you tell me something about your family?

What do your family members do?

Can you tell me something about your club?

How many members are there?

How often do you meet with them?

Can you tell me more about your hobbies?

Why do you like them?

When did you start them?

Can you tell me something about your part-time job?

What is your job?

What do you actually do?

What is your favorite TV program?

What kind of program is it?

Why do you like it?

What is the most interesting news these days?

Why do you think it is interesting?

What do you think about it?

4 . Free form

Please tell me about your life at college.

Do you talk to teachers ? --What are the teachers like ?

Do you ever eat lunch at the dining hall--What is the dining room like ?

Do you go to the library ? ----What is the library like ?

5 . Winding up

4) Dialogue/ Semi-Multilogue (approximately five people in each group) (dialogue in the sense that two people are conversing with each other in each individual part, and semi-multilogue in the sense that more than two people are present in the session) : Group Interview (GI) Test

A classroom teacher is the interlocutor and assessor of this activity.

Topics for Group Interview

1 . What was the most thrilling experience in your life ?

2 . What was the happiest experience in your life ?

3 . What was the most exciting experience in your life ?

4 . What is the most irritating thing which has happened to you recently ?

5 . What was the biggest mistake you have made in your life ?

5) Small Size Multilogue (approximately five people in each group) : Small Group Discussion (SGD) Test

A classroom teacher initiates the discussion by giving students a topic, and has them discuss the topic. The difference between group interview and small group discussion is that in the latter students must focus on the discussion on the given topic so that they are ready to give their opinions when their turn comes.

Topics for Small Group Discussion

1 . Aging Society

2 . Global Warming

3 . Teaching English at Elementary Schools

4 . How to Cope with Mental Stress

6) Large Size Multilogue (more than twenty people in class) : Large Group Discussion (LGD) Test

Each student has to experience two roles at least. One is the leader of the discussion. The other is a participant in the discussion. As the leader they have to mainly do the following things.

(1) to initiate the talk

(2) to give information on the topic

(3) to ask stimulating questions on the topic

(4) to push the participants to express opinions

(5) to keep the discussion on the topic

(6) to make sure no one dominates

Evaluation was conducted through classroom teachers observation based on students' performance. Topics are related to lifestyles, trends, culture, travel, cross-cultural issues as follows:

1 . Limitation of smoking areas

2 . College festival

3 . Music

4 . Eco-driving (Using low pollution vehicles)

5 . Computer Games

6 . Critical thinking in Education

7 . Health Problems

8 . Cultural Awareness

9 . Language Learning using computers

10. Portable Phones

3 . 2. Procedure of the Research

1) subjects: 46 Japanese college students

2)raters:two classroom teachers (one native speaker of English, one non-native speaker of English)

3)The SO, TMS tests were conducted in the language laboratory and recorded on tape.

4)FFI, GI and SGD were conducted once in a classroom situation by each classroom teacher.

5)LGD was the whole term course in which students performances were observed by the native speaker of English.

6)The rating scale was a four-point scale(1:poor 4:good).

4. Results and Discussion

Abbreviations used

SO:Speech Making Overall(SO)(One Way Monologue)

TMS:Tape Mediated Sociolinguistic(TMS) Test(Monologue/Semi-Dialogue)

FFI:Face to Face Interview(FFI) Test(Dialogue)

GI:Group Interview(GI) Test(Dialogue/ Semi-Multilogue)

SGD:Small Group Discussion(SGD)Test (Small Size Multilogue)

LGD:Large Group Discussion(LGD)Test (Large Size Multilogue)

Table 1 shows the inter-rater reliability in each test, except LGD where students' performances were evaluated by a native speaker of English only.

Table 2 demonstrates that the mean score of TMS is the highest followed by GI, SO, SGD, FFI and LGD. TMS demonstrates the highest mean score with the smallest standard deviation. One reason may be that in this type of test, students only have to express the remembered idiomatic or collocation responses in each

Table 1 Inter-rater reliability between two classroom teachers in each test

SO	0.7**
TMS	0.68**
FFI	0.67**
GI	0.61**
SGD	0.62**

**P<.01

Note :
As mentioned earlier, the LGD test was evaluated only by a native speaker of English at the end of the whole term course.

Table 2 Descriptive Statistics of Six Tests Results

	SO	TMS	FFI	GI	SGD	LGD
N	46	46	46	46	46	46
Mean	3	3.17	2.65	3.15	2.87	2.48
S.D.	0.69	0.52	0.76	0.83	0.82	1.1

Table 3 Distibution of Frequencies of Studen's Test Scores (1 - 4 point scale)

		SO	TMS	FFI	GI	SGD	LGD
poor	1	1	0	2	2	2	12
	2	8	3	18	7	13	10
	3	27	32	20	19	20	14
good	4	10	11	6	18	11	10
total		46	46	46	46	46	46

context with no reciprocity, even though this can be called a simulated oral proficiency interview test which involves two people. In other words, the more appropriate phrases or sentences students remember , the better scores they obtain in this type of test.

The mean score of GI is much higher than that of SGD and LGD, although all of them need at least two or more people. One possible explanation for this

result is that in a GI type test, one student can have some preparation time to think about his/her answer while other students are answering the question, especially when the same topic is used. Also, one student can imitate the answering pattern or good phrases of other students who responded earlier. Furthermore, in the GI test, students become less nervous because of the preparation time.

The fact that LGD has the lower mean score with the highest standard deviation can suggest that students discussion ability varies greatly. In other words, there are many students who are not good at large group discussion, although in this kind of test, other factors such as psychological factors or personality factors may play an influential role in the students performance, as we teachers already know from our teaching experience. This is supported by the frequency results in Table 3, too.

Roughly speaking, except GI, there seems to be a difference between the test where one person is involved and the test where more than two persons are involved.

Table 4 Correlation Coefficients of Six Tests (n=46)

	SO	TMS	FFI	GI	SGD	LGD
SO						
TMS	0.42**					
FFI	0.37*	0.21				
GI	0.45**	0.24	0.5**			
SGD	0.42**	0.15	0.66**	0.57**		
LGD	0.17	0.16	0.27	0.35*	0.48**	

Note:**p<.01,*p<.05

Table 4 demonstrates the correlations of six tests. TMS shows low correlations with all the other tests except with SO. One reason may be that in SO and TMS, the number of people involved is only one, while the other tests need at least two people. Another possible reason is that neither one requires listening

ability. The difference between SO and TMS may be that in SO , a student has to demonstrate logically connected sentences. In other words, their discourse level ability is examined. On the other hand, in TMS, a student can focus only on the short idiomatic phrases appropriate in each individual context, and he/she has only to express his/her remembered sentences. Accordingly, the main difference between SO and TMS appears, even if they can be categorized in monologue type in the tape mediated speaking test.

SO results are somewhat correlated with FFI, GI and SGD. This is probably because all of these tests demand some logically longer connected sentences in the responses.

LGD , although it has a somewhat high correlation with SGD, shows low correlations with all the other tests such as SO, TMS, FFI and GI because of the nature of the involvement of the participants in it.

When we look at the 3 highest correlations in the results, FFI, GI and SGD demonstrate the highest correlations with each other than do the others. These three (dialogues and small size multilogue) have something in common as a linguistic activity.

Table 5 Factor Analtsis
(Principal Axis Factoring with unrotated solution)

	Factor1	Factor2
SGD	.89153	-.36083
GI	.68908	.02713
FFI	.68561	-.10110
SO	.62912	.46400
LGD	.45521	-.14649
TMS	.35850	.41027
Eigenvalue	2.47	0.55
Explained Variance	41.2%	9.1%

Table 5 shows the results of factor analysis (Principal Axis Factoring with unrotated solution). Factor 1, which is common to all the tests, is a main factor. This factor may be related to the basic linguistic speaking ability (dealing with grammar, vocabulary, pronunciation etc.), although TMS is a little away from the other four tests because of the nature of the test, as mentioned above. TMS is a type of tape mediated simulated speaking test where students have to give their responses on tape as if they were speaking to a person in a virtual situation. They only have to give short appropriate sentences, in other words, the remembered collocation type sentences. As a whole, this main factor is generalizable basic linguistic ability which is required in any of these speaking tests.

Factor 2 is the one related to the number of people involved in the test. Basically speaking, the more people the test requires, the more negative factor loading it has. In other words, the fewer the test needs, the more positive factor loading it has. Therefore, SO and TMS give more contribution to the factor positively. LGD, although it is giving a negative factor loading, has lower factor loading than SGD. This may be caused by other factors such as the nature of the test.

5. Conclusions and Implications

The following conclusions can be drawn:

1) We were able to find one main general factor which is the generalizable part of speaking ability common to all the tests. And this ability can be measured in terms of basic language aspects such as grammar, vocabulary, pronunciation etc.

2) We were able to obtain a specific function of speaking ability caused by the number of people involved in the test. There seems to be a difference in

students' oral performances between the monologue test and the dialogue or the multilogue test because of the difference of the number of people involved. Another possible cause might be whether listening ability is required or not in the test when we think about the nature of the monologue test where listening is basically omitted.

3) Even in the dialogue and multilogue test, LGD(large group discussion) requires special skills which make the large group discussion different from the small group discussion strategies. Although the construct of the discussion ability is not yet clearly known, this ability should be given some thought from the viewpoint of not only teaching English but also testing English ability, because this kind of ability is most required at the international business and communication level.

The results, as a whole, demonstrate that some students who are good at monologue type tests, even dialogue type tests cannot perform satisfactorily in the discussion test/ activities. This may be caused by the lack of discussion skill in Japanese, by the classroom circumstances where students are hesitant to speak out, or their personalities. Needless to say, it is worthwhile not only to investigate the cause of the students performance differences but also to give a variety of tests for communication ability since they might give positive washback to students and urge them to realize the framework of communication. It also leads the students to develop communication-oriented learning habits.

Since these tests seem to enjoy high face validity in the eyes of both teachers and students, the expected washback will be a solid basis for further assessment of communication ability.

Future research should be carried out by consideration of the following:

1) What is the construct of English speaking ability ?

2) What is the construct of practical basic speaking ability common to all the tests ?

3) How can we effectively and practically evaluate students' large group discussion ability without conducting a long-time in-class observation ?

Note: A version of this paper was presented at the 20th Language Testing Research Colloquium (LTRC), Monterey California, March 9-12, 1998.

Bibliography

Crystal, D. (1995). The Cambridge Encyclopedia of the English Language. Cambridge: CUP.

Hancock, C. (1994). Teaching, testing and assessing: making the connection. Lincolnwood, IL: National Textbook.

Hilsdon. J. (1991). The Group Oral Exam. Advantages and Limitations. In J. C. Alderson and B. North (eds). Language Testing in the 1990s. London: Macmillan. (pp. 189-197).

Jordan, R. R. (1997). English for Academic Purposes. Cambridge: CUP.

Klein-Braley, C (1991). Asking Stupid Questions...: Testing Language Proficiency in the Context of Research Studies. In K. de Bot, R. B. Ginsberg, C. Kramsch (eds). Foreign Language Research in Cross-Cultural Perspective. Amsterdam: John Benjamins. (pp. 73-94).

McNamara, T. (1996). Measuring Second Language Performance. London: Longman.

McNamara, T. (1997). 'Interaction' in second language performance assessment: Whose performance ? Applied Linguistics, 18, 4 , 446-466.

Manidis, M & P. Prescott. (1994). Assessing Oral Language Proficiency. Sydney: National Centre for English Language Teaching and Research; Macquarie University.

Reves, T. (1991). From Testing Research to Educational Policy: A Comprehensive Test of Oral Proficiency. In J.C. Alderson and B. North (eds). Language Testing in the 1990s. London: Macmillan. (pp. 178-188).

Shohamy, E. (1983). The stability of the oral proficiency trait on the oral interview speaking tests. Language Learning 33, 527-39.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. Language Testing 11, 99-124.

Shohamy, E. (1997). Testing methods, testing consequences: are they ethical ? Are they fair? Language Testing 14, 340-349.